Homework 4 Solution

Sta113, ISDS

March 3, 2003

Total 45 points.

0 (5 points)

In this simulation, we first study the sampling distribution of sample mean of 5 observations (that is, m=5) from Weibull distribution with parameters a=2 and b=5. We use Matlab to generate a 1000 by 5 matrix with each element is a random observation from the Weibull distribution. Then we calculate the mean cross row of the matrix and obtain a vector with length=1000 where each element of the vector is the sample mean of 5 observations. Finally, we use the Matlab function sdist to draw the histogram for the 1000 sample means. We repeat this procedure for sample sizes m=25, 50 and 100. We observe that the histograms all are mound-shaped and tend to cluster about the mean of Weibull(2,5), which is equal to 0.7993. Furthermore, as the sample size m increases, there is less variation in the sampling distribution, that is, the histogram is more sharply centered at the mean.

By Central Limiting Theorem, we know that when the sample size m is large enough, the sampling distribution of the sample mean can be approximated by a normal distribution with mean equal to the population mean and variance is equal to the variance of the population distribution divided by m. To check this, we superpose a normal density curve on each figure with appropriate mean and variance. For example, when the sample size m=5, we plot the normal density curve with mean=0.7993 and variance =0.0067. The figures show that when the sample size m increases, the shape of the sampling distribution tends toward the shape of the normal distribution (symmetric and mound-shaped).

1 (6.52, 5 points)

- a.
$$P(x = .3) = p(.3, 0) + p(.3, 10) + p(.3, 20) + p(.3, 30)$$
$$= .07 + .05 + .03 + 0.2 = .17$$

- b.
$$P(y = 20|x = .3) = \frac{p(.3, 20)}{p_1(.3)} = \frac{.03}{.17} = .1765$$

- c. Using the formulas presented in this chapter:
$E(x) = .207$
$E(y) = 10.0$
$E(xy) = 1.76$

$Cov(xy) = E(xy) - E(x)E(y) = 1.76 - (.207)(10) = -.31$
$\Rightarrow x$ and y are correlated.

- d. Since x and y are correlated, they cannot be independent.

2 (7.32, 5 points)

The number of defectives, y, in the sample of 200 has a binomial distribution with $n = 200$ and $p = .08$. The lot will be rejected if more than 6% (12/200) of the sample proves to be defective.

$P(\text{lot rejected}) = P(y > 12) \approx P\left(z > \frac{(12+.5)-np}{\sqrt{npq}}\right)$
$= P\left(z > \frac{12.5-200(.08)}{\sqrt{200(.08)(.92)}}\right) = P(z > -.091)$
$= .5 + P(-.091 < z < 0) = .5 + .3186 = .8186$

3 (5 points)

The marginal distributions for x and y are the same:

$$f(x) = \begin{cases} \frac{1}{4}, & x = -1 \\ \frac{1}{2}, & x = 0 \\ \frac{1}{4}, & x = 1 \\ 0, & \text{otherwise} \end{cases}$$

$E(x) = -1(1/4) + 0(1/2) + 1(1/4) = 0$
$E(xy) = 1/4[(-1)(0) + (0)(-1) + (0)(1) + (1)(0)] = 1/4(0 + 0 + 0 + 0) = 0$

$Cov(xy) = E(xy) - E(x)E(y) = 0 - 0 = 0$
If the covariance is 0 then the correlation is also 0.

To show that they are not independent consider $p(-1, 0) = 1/4 \neq 1/8 = p(-1)p(0)$.

4 (10 points)

- a. The joint pdf of x and y is given by

$$f(x, y) = \begin{cases} 1, & 5 \leq x, y \leq 6 \\ 0, & \text{otherwise} \end{cases}$$

- b.

$$\begin{aligned} P(5.25 \leq x, y \leq 5.75) &= \int_{5.25}^{5.75} \int_{5.25}^{5.75} dx dy \\ &= \frac{1}{4} \end{aligned}$$

- c. Define the region $A = \{(x, y) : 5 \leq x, y \leq 6, |x - y| \leq 1/6\}$. Then the probability is equal to

$$\iint_A dx dy = \text{Area}(A) = 11/36.$$

5 (5 points)

- a. First calculate the marginal density for x,

$$f_x(x) = \int_0^\infty f(x,y)dy = e^{-x}.$$

The probability that the lifetime x exceeds 3 is given by

$$P(x \geq 3) = \int_3^\infty e^{-x}dx = e^{-3}.$$

- b. the marginal density for y is equal to

$$\begin{aligned} f_y(y) &= \int_0^\infty xe^{-x(1+y)}dx \\ &= \frac{\Gamma(2)}{(1+y)^2}\int_0^\infty \frac{(1+y)^2}{\Gamma(2)}xe^{-x(1+y)}dx \end{aligned}$$

Notice that the function inside the integral is the density function for Gamma distribution with parameter $\alpha = 2$ and $\beta = 1/(1+y)$. So

$$f_y(y) = \frac{1}{(1+y)^2},$$

where we use the fact that $\Gamma(2) = 1$.

Since $f(x,y) \neq f_x(x)f_y(y)$, the random variables x and y are NOT independent.

6 (10 points)

- a. By the CDF method, we first calculate the probability of $t \leq t_0$.

$$\begin{aligned} P(t \leq t_0) &= P(x_1 \leq t_0 \text{ AND } x_2 \leq t_0) \\ &= P(x_1 \leq t_0)P(x_2 \leq t_0) \quad \text{Independence} \\ &= \left[\int_0^{t_0} \frac{1}{\beta_1}e^{-x_1/\beta_1}dx_1\right]\left[\int_0^{t_0} \frac{1}{\beta_2}e^{-x_2/\beta_2}dx_2\right] \\ &= (1 - e^{-t_0/\beta_1})(1 - e^{-t_0/\beta_2}) \end{aligned}$$

So the CDF of t is equal to

$$F(t) = (1 - e^{-t/\beta_1})(1 - e^{-t/\beta_2}).$$

Differentiating $F(t)$ with respect to t, we obtain the density function for t.

$$f_t(t) = \frac{1}{\beta_1}e^{-t/\beta_1} + \frac{1}{\beta_2}e^{-t/\beta_2} - \left(\frac{1}{\beta_1} + \frac{1}{\beta_2}\right)e^{-t(1/\beta_1+1/\beta_2)}$$

- b.

$$P(t \geq 6) = \int_6^\infty f_t(t)dt = 1 - F_t(6) = 1 - (1-e^{-3})(1-e^{-4}) = 1 - (.9502)(.9816) = .06719$$

$$(= e^{-3} + e^{-4} - e^{-7} \text{ from exam 1})$$